

A Data-driven Approach to Actuator and Sensor Fault Detection, Isolation and Estimation in Discrete-Time Linear Systems

E. Naderi and K. Khorasani ^a

^a*Department of Electrical and Computer Engineering, Concordia University, Montreal, Quebec, Canada*

Abstract

In this work, we propose explicit state-space based fault detection, isolation and estimation filters that are data-driven and are directly identified and constructed from only the system input-output (I/O) measurements and through estimating the system Markov parameters. The proposed methodology does not involve a reduction step and does not require identification of the system extended observability matrix or its left null space. The performance of our proposed filters is directly connected to and linearly dependent on the errors in the Markov parameters identification process. The estimation filters operate with a subset of the system I/O data that is selected by the designer. It is shown that the proposed filters provide asymptotically unbiased estimates by invoking low order filters as long as the selected subsystem has a stable inverse. We have derived the estimation error dynamics in terms of the Markov parameters identification errors and have shown that they can be directly synthesized from the healthy system I/O data. Consequently, the estimation errors can be effectively compensated for. Finally, we have provided several illustrative case study simulations that demonstrate and confirm the merits of our proposed schemes as compared to methodologies that are available in the literature.

Key words: Data-driven methodology; Actuator and sensor fault diagnosis; Fault detection, isolation, and estimation; Linear discrete-time systems.

1 Introduction

Since the concept of autonomous fault diagnosis, also known as fault detection and isolation (FDI), has been introduced by Beard [1], it has received enormous an attraction in the literature. Some excellent surveys have been published that summarize the extensive literature on fault diagnosis of dynamical systems[2,3]. Two main fault diagnosis categories are model-based and data-driven approaches. The available model-based approaches are reviewed and described in [4].

As engineering systems evolve, it is less likely that engineers have a detailed and accurate mathematical description of the dynamical systems they work with. On the other hand, advances in sensing and data acquisition systems can provide a large volume of raw data for most engineering applications. Consequently, one can find a trend towards data-driven based approaches in many disciplines and problems, including fault diagnosis.

The term ‘data-driven’ covers a wide range of techniques in the literature. Some of the most important strategies

are neural networks [5], fuzzy logic [6], and hybrid approaches [7]. In addition to artificial intelligence based methods, some efforts have been made that are aimed at extending the rich model-based fault diagnosis techniques to data-driven based approaches.

A trivial solution will be the one where one can first identify a mathematical dynamical model of the system from the available data, and then by using the resulting explicit model one then implements and designs conventional model-based schemes. However, this approach suffers from the subsequent errors that are introduced in the system identification process and that may ultimately aggravate the FDI scheme design process errors which can result in a totally unreliable fault diagnosis scheme.

In recent years, a new paradigm has emerged in the literature that aims at direct and explicit construction of the FDI schemes from the available system input-output (I/O) data [8,9,10]. Subspace-based data-driven fault detection and isolation methods [11,12] represent as one of the main approaches that have been reviewed in [11]. These methods are developed based on identifying the left null space of the system extended observability matrix using the I/O data. An estimate of the system

Email address: kash@ece.concordia.ca (E. Naderi and K. Khorasani).

order and an orthogonal basis for the system extended observability matrix - or its left null space - are obtained via the SVD decomposition of a particular data matrix that is constructed from the system I/O data. This process is known as the reduction step.

Essentially, in the reduction step it is assumed that the number of the first set of significantly nonzero singular values and the associated directions provide an estimate of the system order and a basis for the extended observability matrix. However, in most cases, this process leads to erroneous results due to the fact that the truncation point for neglecting small singular values, as being insignificant, is not obvious a trivial and is subjective and problem dependent.

Consequently, an erroneous system order and basis for the extended observability matrix - or its left null space - can be obtained. This error manifests itself in the fault diagnosis scheme performance in a nonlinear manner. In other words, the performance representation of the FDI scheme is not a linear function of the gap between the estimated system order and the system extended observability matrix and the actual ones. Due to these drawbacks, other works that have appeared in the literature are mainly concerned with only the fault estimation problem in which the main objective is to eliminate and remove the above reduction step.

Dong and his colleagues [13] have developed a fault detection scheme that can be directly synthesized from the system I/O data without involving the reduction step. The detection filter is in fact a high order FIR filter parameterized by the system Markov parameters. The extension of this work to the fault isolation task is not trivial and straightforward. It can be performed by obtaining a projection vector that is computed through the SVD decomposition of a transfer matrix parameterized by the Markov parameters estimation errors [14]. However, the Markov parameters estimation errors are not generally available. Therefore, the authors in [14] have managed to synthesize this matrix from the I/O data. The order of the isolation filters can be as large as 30. Dong and Verhaegen [15] used the same strategy for direct construction of the fault estimation filter. The underlying assumption is that the system should have a stable inverse. It will be asymptotically unbiased if the FIR filter order tends to *infinity*.

Wan and his colleagues [16] have reasoned in their recent work that the method of [15] cannot be applied to certain open-loop systems. Moreover, it does not compensate for the estimation errors. Consequently, Wan and his colleagues have proposed offline and online algorithms for compensating for the estimation errors. Yet, it suffers from *two* major drawbacks. First, the

estimation is asymptotically unbiased if the filter order tends to infinity. Second, the computational time per sample for the online optimization algorithm - which is the one that yields an almost unbiased results among the others proposed - is significantly high as compared to the offline methods in [16].

In this work, to overcome the above drawbacks and limitations, we have proposed fault detection, isolation and estimation filters that are constructed directly in the state-space representation form from and using only the available system I/O data. Our proposed schemes only require identification of the system Markov parameters that are achieved by using conventional methods, such as correlation analysis [17] or subspace methods [18,19,20,21] from the healthy I/O data.

Our method does *not* involve the reduction step or equivalent forms of the extended observability matrix. Therefore, the estimation error is linearly dependent on the Markov parameter estimation errors. This step is already addressed in the literature as reviewed above. However, it turns out that our state-space based approach can address several important difficulties that are associated with the currently available works in the literature. First, our proposed identification and isolation filters are conveniently configured for the isolation task of both single *as well as* concurrent faults through constructing filter banks.

An important feature of our proposed state-space based method is that estimation will be achieved asymptotically unbiased by a filter order *as low as* the maximum of the system relative degree and the system observability index. Both of these parameters are bounded by the system order. Moreover, it does not necessarily require the condition of having an entire stable inverse system. The flexibility of our proposed scheme allows arbitrary selection of the subsystems for achieving the fault isolation or for performing the fault estimation tasks.

In other words, one can select a different subsystem if an actuator fault estimation is blocked due to unstable inversion of a specific subsystem. Finally, the state-space based approach allows one to implement a simple and yet effective procedure for compensating for the estimation errors. Towards this end, in this work we derive the estimation error dynamics and show that it can be directly identified from the healthy system I/O data. We will demonstrate through comprehensive simulation studies the effectiveness of our proposed error compensating producer.

The contributions of this paper can now be summarized

as follows:

- (1) A general fault detection and isolation filter for both actuator and sensor faults is developed and directly constructed from only the available system I/O data in the state-space form in a manner that does not involve a reduction step. Moreover, our approach does not require an *a priori* knowledge of the system order. The proposed fault detection and isolation filters can be conveniently configured for both single and concurrent fault detection and isolation tasks by using a subset of the I/O data.
- (2) A fault estimation scheme for both actuator and sensor faults (single and concurrent) is developed and directly constructed from the available system I/O data in the state-space form in a manner that does not involve a reduction step. The proposed estimation filter is asymptotically unbiased having an order as small as the maximum of the observability index and the system relative degree.
- (3) A new offline procedure for *tuning* the estimation filters are proposed to compensate for errors that are caused by the Markov parameters estimation uncertainties.

The outline of the remainder of the paper is as follow. The preliminaries, problem definition and assumptions are provided in Section 2. In Section 3, we discuss the theoretical aspects of our proposed fault estimation scheme. We present the development and design of data-driven fault detection and isolation filters in Section 4. Next, we propose a data-driven fault estimation filter for both the actuators and sensors as well as a tuning procedure is introduced and developed in Section 5. Finally, we provide a number of illustrative simulation results in Section 6.

2 Preliminaries

Consider the following discrete-time linear system \mathbf{S} ,

$$\mathbf{S} : \begin{cases} x(k+1) = Ax(k) + Bu(k) + w(k) \\ y(k) = Cx(k) + v(k) \end{cases} \quad (1)$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, and $y \in \mathbb{R}^l$. Moreover, $w(k) \in \mathbb{R}^n$ and $v(k) \in \mathbb{R}^l$ are white noise having zero mean and covariance matrices:

$$\mathbf{E} \left[\begin{bmatrix} w_i \\ v_i \end{bmatrix} \begin{bmatrix} w_j^T & v_j^T \end{bmatrix} \right] = \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \delta_{i,j} \quad (2)$$

We model a given actuator or a sensor fault through additive terms that are injected in the system \mathbf{S} as follows,

$$\mathbf{S}_f : \begin{cases} x(k+1) = Ax(k) + Bu(k) + Bf^a(k) + w(k) \\ y(k) = Cx(k) + f^s(k) + v(k) \end{cases} \quad (3)$$

where $f^a(k) \in \mathbb{R}^m$ and $f^s(k) \in \mathbb{R}^l$ represent the actuator and sensor faults, respectively. These faults are commonly known as *additive faults*.

Remark 1 *The actuator and sensor faults are traditionally modeled in various manners in the literature. For instance, either as additive faults or multiplicative faults. The proper choice depends on the actual characteristics of the fault. Typically, sensor bias, actuator bias and actuator loss of effectiveness are considered as additive faults. Multiplicative fault models are more suitable for representing changes in the system dynamic parameters such as gains and time constants [22].*

Problem Statement: The problem considered in this work deals with developing and designing fault detection, isolation and estimation schemes for *both* sensor and actuator faults under the following assumptions.

Assumption 1: The system \mathbf{S} is stable and observable.

Assumption 2: The system matrices and the system order are not known *a priori*.

Assumption 3: A sequence of healthy measured system I/O data, namely $u(k)$ and $y(k)$, for $k = 1, \dots, T$, are available and the input $u(k)$ satisfies the persistently exciting (PE) condition [17].

Assumption 4: The faults in the system \mathbf{S}_f are *detectable* and *isolable*, as comprehensively discussed in [23].

The above assumptions are required in all the lemmas and theorems provided in the paper, however they are not explicitly stated in lemmas and theorems statements for sake of brevity.

Identification of the Markov Parameters: We define the set $\{H_0, H_1, H_2, \dots\}$, where $H_\beta = CA^\beta B$ is known as the Markov parameter. If $u(k)$ is persistently exciting, then several approaches are available in the literature to directly identify the Markov parameters from the I/O data $u(k)$ and $y(k)$ ([15,17]). Specifically, we use *correlation analysis* [17] for accomplishing the Markov parameters estimation task. The estimated Markov parameters are denoted by \hat{H}_β in our subsequent derivations.

Notation: We will subsequently use an equivalent form of the system \mathbf{S} as follows,

$$\mathbf{S} : \begin{cases} x(k-i+1) = Ax(k-i) + B\mathbf{I}_l^m \mathbf{U}_i(k-i) + w(k-i) \\ \mathbf{Y}(k-i) = \mathbf{C}x(k-i) + \mathbf{D}\mathbf{U}(k-i) \\ + \mathbf{E}\mathbf{W}(k-i) + \mathbf{V}(k-i) \end{cases} \quad (4)$$

where,

$$\mathbf{C} = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^i \end{pmatrix}; \mathbf{D} = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ H_0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ H_{i-1} & H_{i-2} & \dots & H_0 & 0 \end{pmatrix} \quad (5)$$

$$\mathbf{E} = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ C & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ CA^{i-1} & CA^{i-2} & \dots & C & 0 \end{pmatrix}$$

For any given signal $g(k)$, the following matrices are defined,

$$\mathbf{G}(k-i) = \begin{bmatrix} g(k-i) \\ g(k-i+1) \\ \vdots \\ g(k) \end{bmatrix}; \mathbf{G}_+(k-i) = \begin{bmatrix} g(k-i) \\ g(k-i+1) \\ \vdots \\ g(k+1) \end{bmatrix} \quad (6)$$

We extensively use the notation \mathbf{I}_α^γ which is defined as,

$$\mathbf{I}_\gamma^\alpha = \begin{bmatrix} \mathbf{I}_{\alpha \times \alpha} & \mathbf{0}_{\alpha \times (i\gamma - \alpha)} \end{bmatrix} \quad (7)$$

Moreover, we also define,

$$\mathbf{D}_+ = \begin{pmatrix} H_0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ H_i & H_{i-1} & \dots & H_0 & 0 \end{pmatrix}; \mathcal{D}_+ = \begin{pmatrix} H_0 \\ \vdots \\ H_i \end{pmatrix} \quad (8)$$

$$\mathbf{C}_+ = \begin{bmatrix} (CA)^T & \dots & (CA^{i+1})^T \end{bmatrix}^T$$

$$\mathbf{G}_{i,j}(k-i) = \begin{pmatrix} g(k-i) & g(k-i+1) & \dots & g(k-i+j) \\ g(k-i+1) & g(k-i+2) & \dots & g(k-i+j+1) \\ \vdots & \vdots & \vdots & \vdots \\ g(k) & g(k+1) & \dots & g(k+j) \end{pmatrix} \quad (9)$$

The matrices $\hat{\mathbf{D}}$, $\hat{\mathbf{D}}_+$, $\hat{\mathcal{D}}$ and $\hat{\mathcal{D}}_+$ are constructed similar to \mathbf{D} , \mathbf{D}_+ , \mathcal{D} and \mathcal{D}_+ where the actual Markov parameters H_β are replaced by their estimates \hat{H}_β .

We define two sets p and q that contain a selection of integer numbers from 1 to l and from 1 to m , respectively. The parameters n_p and n_q denote the number of elements in the sets p and q , respectively. Both p and q can be empty sets denoted by $p = \{\emptyset\}$ and $q = \{\emptyset\}$. We denote by kp (or kq), $k \in \mathbb{N}$, as a set that contains all the elements of p (or q) multiplied by k . The notation $\sim q$ (or $\sim p$) denotes the set that contains the integers from 1 to m (or 1 to l) that are not included in q (or p). For example, for a given Markov parameter matrix $H_0 \in \mathbb{R}^{5 \times 4}$, a typical p can be taken as $p = \{2, 4\}$. Moreover, $n_p = 2$, $3p = \{6, 12\}$ and $\sim p = \{1, 3, 5\}$. If $p = \{\emptyset\}$, then $\sim p = \{1, 2, 3, 4, 5\}$. The matrix \mathbf{O}_{p-}^{q-} is obtained by deleting the columns $q, 2q, \dots, iq$ and rows $p, 2p, \dots, ip$ of \mathbf{O} , respectively. The matrix \mathbf{O}^{q+} and \mathbf{O}_{p+} are defined as matrices that only contain the columns $q, 2q, \dots, iq$ and the rows $p, 2p, \dots, ip$ of \mathbf{O} , respectively. The vector \mathbf{P}^{p-} is obtained by deleting the rows $p, 2p, \dots, ip$ of \mathbf{P} . Finally, \mathbf{P}^{p+} only contains the rows $p, 2p, \dots, ip$ of \mathbf{P} . Similar notations are defined for \mathbf{P}^{q+} and \mathbf{P}^{q-} . The signs \dagger , \mathcal{N} and $\mathbb{E}\{\cdot\}$ denote the Moore-Penrose pseudo inverse, null space and the expectation operator. The matrix $\mathbf{O}(\alpha : \beta, \gamma : \theta)$ denotes a matrix that is constructed from an original matrix \mathbf{O} by only containing the rows α to β and the columns γ to θ . If α and β (or γ and θ) are not specified, then it implies that we are dealing with all the rows (or columns) of \mathbf{O} .

Remark 2 The parameters p and q are defined in order to specify the set of I/O data that is to be fed to a fault isolation or estimation filter. For example, for a given Markov parameter matrix $H_0 \in \mathbb{R}^{5 \times 4}$, one may desire to design a filter that operates with information from input channels $\{3, 4\}$ and measurement channels $\{1, 2, 3\}$. Then, one should set $q = \{1, 2\}$ and $p = \{4, 5\}$. The above notation is critical for the task of fault isolation where one requires to construct a bank of filters each of which operates with a different set of inputs and outputs data.

3 Proposed Fault Estimation Scheme using Exact Markov Parameters and Observability Matrix

In this section, we start by assuming availability of the exact Markov parameters and the extended observability matrix to introduce the basic concepts we utilize in this work. These assumptions will be relaxed and removed in the subsequent Sections 4, 5 and 6.

Let us consider a signal $\eta(k)$ that is governed by the following dynamics and stimulated by the information from the sensors $\sim p$ and actuators $\sim q$, that is

$$\eta(k+1) = A_r \eta(k) + B_r \mathbf{U}^{q-}(k-i) + L_r \mathbf{Y}^{p-}(k-i) \quad (10)$$

where $\eta(k) \in \mathbb{R}^{i'}$, $\mathbf{U}^{q-}(k-i) \in \mathbb{R}^{im'}$ and $\mathbf{Y}^{p-} \in \mathbb{R}^{il'}$, where $l' = (l - n_p)$ and $m' = (m - n_q)$. Our goal is to determine the *unknown* matrices A_r , B_r and L_r such that for the healthy system \mathbf{S} given by (4), we have

$$\mathbb{E}(e(k)) = \mathbb{E}(\eta(k) - \mathbf{T}x(k-i)) \rightarrow 0 \text{ as } k \rightarrow \infty \quad (11)$$

where $\mathbf{T} \in \mathbb{R}^{l' \times n}$ denotes a full column rank matrix. The error dynamics associated with $e(k)$ is therefore given by,

$$\begin{aligned} e(k+1) &= \eta(k+1) - \mathbf{T}x(k-i+1) \\ &= A_r e(k) + (A_r \mathbf{T} - \mathbf{T}A + L_r \mathbf{C}_{p-})x(k-i) \\ &\quad + (B_r + L_r \mathbf{D}_{p-}^{q-} - \mathbf{T}B^{q-} \mathbf{I}_{l'}^{m'})\mathbf{U}^{q-}(k-i) \\ &\quad + (L_r \mathbf{D}_{p-}^{q+} - \mathbf{T}B^{q+} \mathbf{I}_{l'}^{n_q})\mathbf{U}^{q+}(k-i) \\ &\quad + L_r \mathbf{E}_{p-}^{q-} \mathbf{W}^{q-}(k-i) + L_r \mathbf{V}^{p-}(k-i) \end{aligned} \quad (12)$$

which is obtained by substituting $\eta(k+1)$ from equation (10) and $x(k-i+1)$ from equation (4). Condition (11) is now satisfied if and only if (a) A_r is a Hurwitz matrix, (b) $A_r \mathbf{T} - \mathbf{T}A + L_r \mathbf{C}_{p-} = 0$, (c) $B_r + L_r \mathbf{D}_{p-}^{q-} - \mathbf{T}B^{q-} \mathbf{I}_{l'}^{m'} = 0$, and (d) $L_r \mathbf{D}_{p-}^{q+} - \mathbf{T}B^{q+} \mathbf{I}_{l'}^{n_q} = 0$. The above conditions actually correspond to the Luenberger observer equations.

The key concept that is introduced in this paper is that we specifically set,

$$\mathbf{T} = \mathbf{C}_{p-} \quad (13)$$

In other words, we select \mathbf{T} to be equal to the extended observability matrix. Let us now define the matrix \mathbf{M}_{p-} as follows,

$$\mathbf{M}_{p-} = A_r + L_r \quad (14)$$

Given (13) and (14) and in view of the fact that $\mathbf{C}_{p-} B^{q-} = \mathcal{D}_{+,p-}^{q-}$, $\mathbf{C}_{p-} B^{q+} = \mathcal{D}_{+,p-}^{q+}$, and $\mathbf{C}_{p-} A =$

$\mathbf{C}_{+,p-}$, the conditions (a) to (d) can be rewritten as,

$$A_r \text{ is Hurwitz} \quad (15)$$

$$\mathbf{M}_{p-} \mathbf{C}_{p-} = \mathbf{C}_{+,p-} \quad (16)$$

$$L_r \mathbf{D}_{p-}^{q+} - \mathcal{D}_{+,p-}^{q+} \mathbf{I}_{l'}^{n_q} = 0 \quad (17)$$

$$B_r + L_r \mathbf{D}_{p-}^{q-} - \mathcal{D}_{+,p-}^{q-} \mathbf{I}_{l'}^{m'} = 0 \quad (18)$$

Remark 3 Recall that \mathbf{C}_{p-} should be full column rank according to the assignment (13). The matrix \mathbf{C}_{p-} will be full column rank if i is selected to be equal to or greater than the observability index of the pair (C^{p-}, A) , which is denoted by ν_p . Given $i \geq \nu_p$, the matrix \mathbf{M}_{p-} is given by $\mathbf{C}_{+,p-}(\mathbf{C}_{p-})^\dagger + \Theta(\mathbf{I} - \mathbf{C}_{p-}(\mathbf{C}_{p-})^\dagger)$, where Θ is an arbitrary matrix introduced due to the Moore-Penrose pseudo inverse non-unique solution.

Definition 1 The relative degree of the subsystem inputs to the outputs $\sim p$ is defined as the smallest non-negative τ_p such that $H_{i,p-} = 0$ for $i < \tau_p$ and \mathcal{D}_{p-} is full column rank for $i \geq \tau_p$ [16,24].

Remark 4 It is well-known that the system of equations (14) and (15) to (18) has a solution if and only if i) $i \geq \tau_p$ and ii) the subsystem from the inputs q to the outputs $\sim p$ is minimum phase [25]. Particularly, the above equations always have a solution if $q = \{\emptyset\}$ since equation (17) vanishes. Consequently, one can arbitrarily select a Hurwitz A_r and then calculate L_r and B_r from equations (14) and (18). The restriction that is imposed is actually on the subsystems and not the entire system. Therefore, the designer has a freedom to select a different q if the minimum phase condition is not satisfied for the original selection.

Let us assume that it is desired to estimate the faults in the actuators q by using the information from the sensors $\sim p$ and actuators $\sim q$. Then, the fault estimator filter is given by,

$$\begin{cases} \eta(k+1) = A_r \eta(k) + B_r \mathbf{U}^{q-}(k-i) + L_r \mathbf{Y}^{p-}(k-i) \\ \hat{f}^a(k-i) = -\mathbf{I}_m^m \mathbf{D}^\dagger (\eta(k) - \\ \mathbf{Y}^{p-}(k-i) + \mathbf{D}_{p-} \mathbf{U}(k-i)) \end{cases} \quad (19)$$

where $\hat{f}^a(k)$ denotes an estimate of $f^a(k)$. The matrices A_r , B_r and L_r are obtained by solving the equations (14) and (15) to (18).

Theorem 1 Assume that the subsystem from the inputs q to the outputs $\sim p$ is minimum phase, $i \geq \max\{\nu_p, \tau_p\}$ and the sensors $\sim p$ and actuators $\sim q$ are healthy, then the filter dynamics governed by (19) is asymptotically unbiased.

PROOF. The proof is provided in the Appendix A. ■

The above theorem guarantees that our proposed filter will generally have a lower order than the estimation filter that is proposed in [16] which requires $i \geq \nu_p + \tau_p$ or $i \rightarrow \infty$ depending on the transmission zeros of the quadruple $(A, B^{q-}, C_{p-}, D_{p-}^{q-})$.

The case of the sensor fault estimation is slightly different. One can estimate the faults in the sensors p if the sensors $\sim p$ and all the actuators are healthy. The fault estimator is now given by,

$$\begin{cases} \eta(k+1) = A_r \eta(k) + B_r \mathbf{U}(k-i) + L_r^{p\downarrow} \mathbf{Y}(k-i) \\ \hat{f}^s(k-i) = \mathbf{I}_l^l (\eta(k) - \mathbf{Y}^{p-}(k-i) + \mathbf{D}_{p-} \mathbf{U}(k-i)) \end{cases} \quad (20)$$

where $\hat{f}^s(k)$ denotes an estimate of $f^s(k)$, $L_r^{p\downarrow}$ denotes a matrix where its columns $p, 2p, \dots, ip$ are zero. The matrices A_r , B_r and $L_r^{p\downarrow}$ are obtained by solving the equations (14) and (15) to (18) by setting $p = q = \{\emptyset\}$ and by replacing L_r with $L_r^{p\downarrow}$.

The above proposed filter (20) is unbiased as established by the following theorem.

Theorem 2 *Assume that $i \geq \nu_p$ and the sensors $\sim p$ and all the actuators are healthy, then the filter dynamics that is governed by (20) is asymptotically unbiased.*

PROOF. The proof is provided in the Appendix B. ■

Note that the subsystem from the inputs to the outputs $\sim p$ is *not* required to be minimum phase for solving the sensor fault estimation problem. Moreover, it can be theoretically shown that the filter (20) is unbiased by using information from the faulty actuators provided it is modified by the estimations that are provided by the filter (19). However, this coupling will cause significant biases in the data-driven solution since the actuator fault estimation scheme is itself biased.

4 Data-driven Fault Detection and Isolation (FDI) Scheme

In this section, our proposed fault detection and isolation (FDI) filters are now directly constructed from the healthy system I/O data. First, we propose a data-driven estimation of the matrix \mathbf{M}_{p-} and then present the design procedure of the FDI filters.

Remark 5 *Theorems 1 and 2 provide the guidelines for selection of i . The parameter i is bounded by n which is not known a priori. The condition $i \geq \tau_p$ can be easily satisfied by checking the rank of \mathcal{D}_{p-} . However, the parameter ν_p is not known. Therefore, i should be selected sufficiently large that ensures $i \geq \max\{\nu_p, \tau_p\}$.*

4.1 Data-driven Estimation of the Filter Parameters

In order to solve equations (14) and (15) to (18), one requires the Markov parameters and the extended observability matrix. The extended observability matrix is required to obtain \mathbf{M}_{p-} . However, in our subsequent data-driven derivations we will show that an estimate of the matrix \mathbf{M}_{p-} can be directly obtained from the system I/O data *without* applying the reduction step. Consequently, the matrix \mathbf{C}_{p-} or its equivalent forms are not actually required.

The objective of the equation (16) is in fact to enforce,

$$(A_r \mathbf{C}_{p-} - \mathbf{C}_{p-} A + L_r \mathbf{C}_{p-}) x(k-i) \equiv 0 \quad (21)$$

On the other hand, from the measurement equation (4) it follows that,

$$\begin{aligned} \mathbf{C}_{p-} x(k-i) &= \mathbf{Y}^{p-}(k-i) - \mathbf{D}_{p-} \mathbf{U}(k-i) \\ &\quad - \mathbf{E}_{p-} \mathbf{W}(k-i) - \mathbf{V}^{p-}(k-i) \end{aligned} \quad (22)$$

By substituting equation (22) into equation (21) one obtains,

$$\begin{aligned} &\mathbf{M}_{p-} (\mathbf{Y}^{p-}(k-i) - \mathbf{D}_{p-} \mathbf{U}(k-i) - \mathbf{E}_{p-} \mathbf{W}(k-i) \\ &\quad - \mathbf{V}^{p-}(k-i)) - (\mathbf{Y}^{p-}(k-i+1) - \mathbf{D}_{+,p-} \mathbf{U}_+(k-i) \\ &\quad - \mathbf{E}_{+,p-} \mathbf{W}_+(k-i) - \mathbf{V}^{p-}(k-i+1)) \\ &= 0 \end{aligned} \quad (23)$$

where,

$$\mathbf{M}_{p-} \triangleq A_r + L_r \quad (24)$$

Iterating the equation (23) from the time steps $k-i$ to $k-i+j$, where $j \gg i$, yields,

$$\begin{aligned} &\mathbf{M}_{p-} (\Gamma_0^{p-} - \mathbf{E}_{p-} \mathbf{W}_{i,j}(k-i) - \mathbf{V}_{i,j}^{p-}(k-i)) \\ &\quad - (\Gamma_1^{p-} - \mathbf{E}_{+,p-} \mathbf{W}_{(i+1),j}^{p-}(k-i) - \mathbf{V}_{(i+1),j}^{p-}(k-i+1)) = 0 \end{aligned} \quad (25)$$

where $\Gamma_0^{p-} = \mathbf{Y}_{i,j}^{p-}(k-i) - \mathbf{D}_{p-} \mathbf{U}_{i,j}(k-i)$ and $\Gamma_1^{p-} = \mathbf{Y}_{(i+1),j}^{p-}(k-i+1) - \mathbf{D}_{+,p-} \mathbf{U}_{(i+1),j}(k-i)$. Equation (25) forms the basis for our proposed data-driven solution for estimating \mathbf{M}_{p-} . The matrix \mathbf{M}_{p-} minimizes the following cost function,

$$\|\Gamma_1^{p-} - \mathbf{M}_{p-} \Gamma_0^{p-}\|_2$$

We do not have access to the actual values of the Markov parameters. Instead, we construct the matrices $\hat{\Gamma}_0^{p-}$ and $\hat{\Gamma}_1^{p-}$ by using the estimated Markov parameters and the system I/O data (healthy data) as follows,

$$\hat{\Gamma}_0^{p-} = \mathbf{Y}_{i,j}^{p-}(k-i) - \hat{\mathbf{D}}_{p-} \mathbf{U}_{i,j}(k-i) \quad (26)$$

$$\hat{\Gamma}_1^{p-} = \mathbf{Y}_{(i+1),j}^{p-}(k-i+1) - \hat{\mathbf{D}}_{+,p-} \mathbf{U}_{(i+1),j}(k-i) \quad (27)$$

where $\hat{\mathbf{D}}_{p-}$ and $\hat{\mathbf{D}}_{+,p-}$ are constructed similar to \mathbf{D}_{p-} and $\mathbf{D}_{+,p-}$ but instead the estimated Markov parameters are utilized. Therefore, an estimate of \mathbf{M}_{p-} is given by $\hat{\Gamma}_1^{p-}(\hat{\Gamma}_0^{p-})^\dagger + \Theta(\mathbf{I} - \hat{\Gamma}_0^{p-}(\hat{\Gamma}_0^{p-})^\dagger)$, where Θ is an arbitrary matrix. However, the solution will be unique as $j \rightarrow \infty$ as stated in the following lemma.

Lemma 1 *If $j \rightarrow \infty$, then the matrix $\hat{\Gamma}_0^{p-}$ is full row rank and,*

$$\hat{\mathbf{M}}_{p-} = \hat{\Gamma}_1^{p-}(\hat{\Gamma}_0^{p-})^\dagger \quad (28)$$

PROOF. The proof is provided in the Appendix C. ■

The matrix $\hat{\mathbf{M}}_{p-}$ has a particular structure as shown in the following lemma.

Lemma 2 *The matrix $\hat{\mathbf{M}}_{p-}$ has the following structure,*

$$\hat{\mathbf{M}}_{p-} = \begin{bmatrix} 0_{(i-1)l' \times l'} & \mathbf{I}_{(i-1)l' \times (i-1)l'} \\ \mathbf{K}_1 & \mathbf{K}_2 \end{bmatrix} \quad (29)$$

where $\mathbf{K}_1 \in \mathbb{R}^{l' \times l'}$ and $\mathbf{K}_2 \in \mathbb{R}^{l' \times (i-1)l'}$ are nonzero matrices, where $l' = l - n_p$.

PROOF. One can partition $\hat{\Gamma}_0^{p-}$ and $\hat{\Gamma}_1^{p-}$ as follows,

$$\hat{\Gamma}_0^{p-} = \begin{bmatrix} \hat{\Gamma}_{01} \\ \hat{\Gamma}_{02} \end{bmatrix}; \hat{\Gamma}_1^{p-} = \begin{bmatrix} \hat{\Gamma}_{11} \\ \hat{\Gamma}_{12} \end{bmatrix} \quad (30)$$

where $\hat{\Gamma}_{01} \in \mathbb{R}^{l' \times il'}$, $\hat{\Gamma}_{02} \in \mathbb{R}^{(i-1)l' \times il'}$, $\hat{\Gamma}_{11} \in \mathbb{R}^{l' \times il'}$ and $\hat{\Gamma}_{12} \in \mathbb{R}^{(i-1)l' \times il'}$. It follows readily from definitions of $\hat{\Gamma}_0^{p-}$ and $\hat{\Gamma}_1^{p-}$ (equations (26) and (27)) that $\hat{\Gamma}_{11} = \hat{\Gamma}_{02}$. Since $\hat{\Gamma}_0^{p-}$ is full row rank, therefore we have

$$\hat{\Gamma}_{11} = \begin{bmatrix} 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{\Gamma}_{01} \\ \hat{\Gamma}_{02} \end{bmatrix} \quad (31)$$

Moreover, $\hat{\Gamma}_{12}$ can be written as a linear combination of the rows of $\hat{\Gamma}_{01}$ and $\hat{\Gamma}_{02}$ as follows,

$$\hat{\Gamma}_{12} = \begin{bmatrix} \mathbf{K}_1 & \mathbf{K}_2 \end{bmatrix} \begin{bmatrix} \hat{\Gamma}_{01} \\ \hat{\Gamma}_{02} \end{bmatrix} \quad (32)$$

Consequently, we obtain,

$$\hat{\Gamma}_1^{p-} = \begin{bmatrix} 0 & \mathbf{I} \\ \mathbf{K}_1 & \mathbf{K}_2 \end{bmatrix} \hat{\Gamma}_0^{p-} \quad (33)$$

which reveals the general structure of $\hat{\mathbf{M}}_{p-}$ as given by equation (29). ■

Lemma 3 *The matrix $\hat{\mathbf{M}}_{p-}$ is Hurwitz.*

PROOF. The proof is provided in the Appendix D. ■

The above analysis shows that the estimation filter (10), which satisfies the condition (11), can be directly synthesized from the system I/O data *without* requiring any reduction step. The data-driven counterparts of equations (14) and (15) to (18) is now given by,

$$\hat{A}_r \text{ is Hurwitz} \quad (34)$$

$$\hat{A}_r + \hat{L}_r = \hat{\mathbf{M}}_{p-} \quad (35)$$

$$\hat{L}_r \hat{\mathbf{D}}_{p-}^{q+} - \hat{\mathcal{D}}_{+,p-}^{q+} \mathbf{I}_{l'}^{n_q} = 0 \quad (36)$$

$$\hat{B}_r + \hat{L}_r \hat{\mathbf{D}}_{p-}^{q-} - \hat{\mathcal{D}}_+^{q-} \mathbf{I}_{l'}^{m'} = 0 \quad (37)$$

Theorem 3 *Equations (34) to (37) have a solution if and only if $\hat{\mathbf{M}}_{p-} - \hat{\mathcal{D}}_{+,p-}^{q+} \mathbf{I}_{l'}^{n_q} (\hat{\mathbf{D}}_{p-}^{q+})^\dagger - \Theta_1 (\mathbf{I} - \hat{\mathbf{D}}_{p-}^{q+} (\hat{\mathbf{D}}_{p-}^{q+})^\dagger)$ is Hurwitz for an arbitrary matrix $\Theta_1 \in \mathbb{R}^{il' \times il'}$.*

PROOF. Solving equation (36) for \hat{L}_r yields,

$$\hat{L}_r = \hat{\mathcal{D}}_{+,p-}^{q+} \mathbf{I}_{l'}^{n_q} (\hat{\mathbf{D}}_{p-}^{q+})^\dagger + \Theta_1 (\mathbf{I} - \hat{\mathbf{D}}_{p-}^{q+} (\hat{\mathbf{D}}_{p-}^{q+})^\dagger)$$

Substituting the above expression in equation (35) and comparing it with equation (34) concludes the result. ■

The arbitrary matrix Θ_1 in \hat{L}_r should be selected such that equation (34) is satisfied. Note that equation (36) and the free parameter Θ_1 vanish when $q = \{\emptyset\}$. Therefore, one can arbitrarily select a Hurwitz matrix \hat{A}_r and then obtain \hat{L}_r and \hat{B}_r from equations (36) and (37).

4.2 Fault Detection and Isolation Filters

Our proposed residual generator fault detection and isolation filter has the general structure that is as governed by,

$$\begin{cases} \eta(k+1) = \hat{A}_r \eta(k) + \hat{B}_r \mathbf{U}^{q-}(k-i) + \hat{L}_r \mathbf{Y}^{p-}(k-i) \\ \hat{r}(k) = \mathbf{I}_f(\eta(k) - \mathbf{Y}_i^{p-}(k-i) + \hat{\mathbf{D}}_{p-} \mathbf{U}(k-i)) \end{cases} \quad (38)$$

where $\hat{r}(k)$ denotes the residual signal, $\mathbf{I}_f = \mathbf{I}_m^m$ if $q \neq \{\emptyset\}$ and $\mathbf{I}_f = \mathbf{I}$, otherwise. The filter parameters are obtained from equations (34) to (37). The above general structure can be configured for both single or concurrent fault detection, isolation or estimation tasks by invoking different settings for p and q . Specifically, if both p and q are set to be empty sets, then the filter

(38) will be in fact a fault detection filter.

Fault isolation is typically performed via *structured residuals*. In other words, a bank of residual observers is constructed where each filter in the bank is insensitive to a particular fault but sensitive to all the other faults. Therefore, in case of occurrence of a fault, all the filters generate non-zero residuals that exceed their thresholds except for the one filter that can be used for determining the isolated fault. This can be achieved by invoking different settings for p and q for each filter in the bank.

For example, if a single actuator fault isolation scheme for a system with three inputs and four measurements is desired, then a bank that consists of three filters should be constructed. A possible configuration setting for the filters 1, 2 and 3 in the bank is $q = \{1, 2\}$, $q = \{1, 3\}$ and $q = \{2, 3\}$, respectively, and $p = \{\emptyset\}$ for all. Alternatively, one may try the setting $q = \{1\}$, $p = \{3, 4\}$, $q = \{2\}$, $p = \{1, 2, 4\}$, and $q = \{1, 3\}$, $p = \{1, 3\}$ for the filters 1, 2 and 3, respectively. A particular configuration selection depends on the context of the problem and the requirements. Despite the above flexibility, our proposed scheme has a limitation that it cannot handle simultaneous concurrent actuator and sensor faults. This situation differs from the concurrent actuators or concurrent sensors faults which is well managed within our proposed framework.

The residuals that are generated by the filter (38) has an important property that is characterized in the next lemma.

Lemma 4 *Given $q = \{\emptyset\}$ and \hat{A}_r selected to be a diagonal Hurwitz matrix, then the first $(i-1)l'$ rows of $\hat{r}(k)$ converge to zero as $k \rightarrow \infty$ independent of the presence of the faults.*

PROOF. The proof is provided in the Appendix E. ■

Based on the above lemma, the first $(i-1)l'$ rows of $\hat{r}(k)$ do not contain any useful information that would allow a model reduction. The general structure of the residual generator filter for the actuator or sensor fault detection or the sensor fault isolation ($q = \{\emptyset\}$) is then given by,

$$\begin{cases} \eta_r(k+1) = \hat{A}_r \eta_r(k) + \hat{B}_r \mathbf{U}(k-i) + \hat{L}_r \mathbf{Y}^{p-}(k-i) \\ \hat{r}(k) = \eta_r(k) - y^{p-}(k) + \hat{\mathbf{D}}_{p-} \mathbf{U}(k-i) \end{cases} \quad (39)$$

where $\eta_r(k) \in \mathbb{R}^{l'}$, $\hat{A}_r = A_r((i-1)l' + 1 : il', il' + 1 : il')$, $\hat{B}_r = B_r((i-1)l' + 1 : il', :)$, $\hat{L}_r = L_r((i-1)l' + 1 : il', :)$ and $\hat{\mathbf{D}}_{i,p-} = \mathbf{D}_{i,p-}((i-1)l' + 1 : il', :)$.

The same model reduction procedure cannot be applied to the actuator fault isolation filter (i.e. when $q \neq \{\emptyset\}$) since \hat{A}_r that is obtained from equations (34) to (37) is not necessarily diagonal.

4.3 Residual Dynamics in Presence of a Fault

Let us now investigate the corresponding residual dynamics in presence of faults. If $f^a(k)$ and/or $f^s(k)$ are nonzero, then the residual dynamics is given by,

$$\begin{cases} \eta(k+1) = \hat{A}_r \eta(k) + \hat{B}_r (\mathbf{U}^{q-}(k-i) + \mathbf{F}^{a,q-}(k-i)) \\ \quad + \hat{L}_r (\mathbf{Y}^{p-}(k-i) + \mathbf{F}^{s,p-}(k-i)) \\ \hat{r}(k) = \mathbf{I}_f (\eta(k) - \mathbf{Y}_i^{p-}(k-i) - \mathbf{F}^{s,p-}(k-i) \\ \quad + \hat{\mathbf{D}}_{p-} (\mathbf{U}(k-i) + \mathbf{F}^a(k-i))) \end{cases} \quad (40)$$

where $\mathbf{F}^a(k-i)$ and $\mathbf{F}^s(k-i)$ are construed similar to $\mathbf{G}(k-i)$ using the actuator fault and the sensor fault signals, respectively. The vectors $\mathbf{F}^{a,q-}(k-i)$ and $\mathbf{F}^{s,p-}(k-i)$ are then obtained by deleting the rows q, \dots, iq and p, \dots, ip of $\mathbf{F}^a(k-i)$ and $\mathbf{F}^s(k-i)$, respectively. The filter dynamics (40) shows that the residual $\hat{r}(k)$ is clearly affected by the faults except those in the sensors p or the actuators q .

Conventionally, the following decision logic is utilized for performing the fault detection task, namely

$$\begin{cases} \text{If } r_{min} \leq \mathbb{E}\{\|\hat{r}(k)\|\} \leq r_{max} \\ \Rightarrow \text{System is healthy} \\ \text{If } \mathbb{E}\{\|\hat{r}(k)\|\} < r_{min} \text{ or } \mathbb{E}\{\|\hat{r}(k)\|\} > r_{max} \\ \Rightarrow \text{System is faulty} \end{cases} \quad (41)$$

where r_{min} and r_{max} denote the lower and the upper bound thresholds, respectively. The thresholds are selected through conducting comprehensive Monte Carlo simulation runs so that the missed alarms and false alarms are minimized.

A similar structure can be utilized for selecting the fault isolation decision logic. Specifically, a fault is detected and isolated in the actuator q_0 if,

$$\begin{cases} \mathbb{E}\{\|\hat{r}^a(k)\|\} < r_{min} \text{ or } \mathbb{E}\{\|\hat{r}^a(k)\|\} > r_{max} ; \\ a \neq q_0, a = 1, \dots, m, \text{ and} \\ r_{min} \leq \mathbb{E}\{\|\hat{r}^a(k)\|\} \leq r_{max} \end{cases} ; a = q_0. \quad (42)$$

where $\hat{r}^a(k)$ denotes the residual that is obtained by setting $q = \{a\}$.

This completes our proposed solution to the problem of data-driven fault detection and isolation. In the next section, we consider the problem of data-driven fault estimation.

5 The Proposed Fault Estimation Scheme

In many practical control problems, it is crucial to estimate the faults once they are detected and isolated. In this section, we provide a data-driven based methodology for design of fault estimation filters. Our proposed fault estimation scheme can be integrated with the FDI scheme. In other words, the FDI scheme introduced in the previous section can be utilized to distinguish between the healthy actuators and sensors from those where their data are used for fault estimation of faulty actuators and sensors.

In this section, we first propose fault estimation filters. It turns out that these filters are biased due to presence of estimation errors in the Markov parameters and the matrix $\hat{\mathbf{M}}_{p-}$. We then derive the dynamics corresponding to the fault estimation errors and show that it can be directly identified from the healthy system I/O data. Finally, we propose our so-called *tuned fault estimation filters* that are obtained for a reliable and actuator fault estimation by integrating the proposed estimation filters with the identified estimation error dynamics.

5.1 Sensor Fault Estimation Filters

The following filter is now proposed to estimate the faults in the sensors p by using data from the sensors $\sim p$ and all the actuators,

$$\begin{cases} \eta(k+1) = \hat{A}_r \eta(k) + \hat{B}_r \mathbf{U}(k-i) + \hat{L}_r^{p\downarrow} \mathbf{Y}(k-i) \\ \hat{f}^s(k-i) = \mathbf{I}_l^l (\eta(k) - \mathbf{Y}(k-i) + \hat{\mathbf{D}} \mathbf{U}(k-i)) \end{cases} \quad (43)$$

where $\hat{f}^s(k-i)$ denotes an estimate of $f^s(k-i)$ and the filter parameters are obtained from the equations (34) to (37) by setting $p = q = \{\emptyset\}$ and by replacing \hat{L}_r with $\hat{L}_r^{p\downarrow}$.

Clearly, the filter (43) is biased due to presence of estimation errors in the Markov parameters and the matrix $\hat{\mathbf{M}}$. The matrix $\hat{\mathbf{M}}$ is defined to be the same as $\hat{\mathbf{M}}_{p-}$ when $p = \{\emptyset\}$. Let us define the estimation error as $\Delta f^s(k) = f^s(k-i) - \hat{f}^s(k-i)$. Therefore, we have,

$$\begin{aligned} \Delta f^s(k-i) &= \mathbf{I}_l^l (\xi(k) + \Delta \mathbf{D} \mathbf{U}(k-i) \\ &\quad + \mathbf{E} \mathbf{W}(k-i) + \mathbf{V}(k-i)) \end{aligned} \quad (44)$$

where $\xi(k) = \mathbf{C}x(k-i) - \eta(k)$ and $\Delta \mathbf{D} = \mathbf{D} - \hat{\mathbf{D}}$. The dynamics of $\xi(k)$ is now governed by,

$$\begin{aligned} \xi(k+1) &= \hat{A}_r \xi(k) + (\hat{A}_r - \mathbf{M} + \hat{L}_r^{p\downarrow}) \mathbf{C}x(k-i) \\ &\quad + (\hat{B}_r + \hat{L}_r^{p\downarrow} \mathbf{D} - \mathbf{C} \mathbf{B} \mathbf{I}_l^m) \mathbf{U}(k-i) \\ &\quad + \hat{L}_r^{p\downarrow} \mathbf{E} \mathbf{W}(k-i) + \hat{L}_r^{p\downarrow} \mathbf{V}(k-i) \end{aligned} \quad (45)$$

The matrix \mathbf{M} is equal to \mathbf{M}_{p-} when $p = \{\emptyset\}$. We substitute $\mathbf{C}x(k-i)$ by $\mathbf{Y}(k-i) - \mathbf{D} \mathbf{U}(k-i) - \mathbf{F}^s(k-i) - \mathbf{E} \mathbf{W}(k-i) - \mathbf{V}(k-i)$ in the above equation. Rearranging of the right hand side of the above equation after substitution yields the governing dynamics of the fault estimation error as follows,

$$\begin{cases} \xi(k+1) = \hat{A}_r \xi(k) + \Delta_u \mathbf{U}(k-i) + \Delta_y \mathbf{Y}(k-i) - \\ \Delta_y \mathbf{F}^s(k-i) + \mathcal{M}_1 \mathbf{W}(k-i) + \mathcal{M}_2 \mathbf{V}(k-i) \\ \Delta f^s(k-i) = \mathbf{I}_l^l (\xi(k) + \Delta \mathbf{D} \mathbf{U}(k-i) \\ + \mathbf{E} \mathbf{W}(k-i) + \mathbf{V}(k-i)) \end{cases} \quad (46)$$

where $\delta A = \hat{A}_r - \mathbf{M} + \hat{L}_r^{p\downarrow}$, $\delta B = \hat{B}_r + \hat{L}_r^{p\downarrow} \mathbf{D} - \mathbf{C} \mathbf{B} \mathbf{I}_l^m$, $\Delta_u = \delta B - \delta A \mathbf{D}$, $\Delta_y = \delta A$, $\mathcal{M}_1 = \hat{L}_r^{p\downarrow} \mathbf{E} - \delta A \mathbf{E}$ and $\mathcal{M}_2 = \hat{L}_r^{p\downarrow} \delta A$.

Equation (46) clearly shows that the fault estimates are biased. All the parameters in the filter (46) are unknown since their computation requires the exact Markov parameters and the matrix \mathbf{M} . However, we will show that one can actually obtain an estimate of Δ_u , Δ_y and $\Delta \mathbf{D}$ by using the healthy I/O data.

Towards this end, we split the off-line available healthy I/O data into two segments. The first segment is utilized to estimate the system Markov parameters and the matrix $\hat{\mathbf{M}}$. Once the filter (43) is constructed, it is stimulated by the second segment of the I/O data to obtain $\hat{f}^s(k-i)$ using an arbitrary initial condition for $\eta(0)$ in (43). Theoretically, $\hat{f}^s(k-i)$ should be zero corresponding to the second segment of the healthy data, however, it will be biased due to presence of the estimation errors in the Markov parameters and the matrix $\hat{\mathbf{M}}$. The bias is governed and is given by $\Delta f^s(k) = f^s(k-i) - \hat{f}^s(k-i) \equiv -\hat{f}^s(k-i)$, and according to equation (46) is governed by,

$$\begin{cases} \xi(k+1) = \hat{A}_r \xi(k) + \Delta_u \mathbf{U}(k-i) + \Delta_y \mathbf{Y}(k-i) - \\ + \mathcal{M}_1 \mathbf{W}(k-i) + \mathcal{M}_2 \mathbf{V}(k-i) \\ \Delta f^s(k) = \mathbf{I}_l^l (\xi(k) + \Delta \mathbf{D} \mathbf{U}(k-i) \\ + \mathbf{E} \mathbf{W}(k-i) + \mathbf{V}(k-i)) \end{cases} \quad (47)$$

One may consider the filter (47) as a stochastic LTI system that is described by the quadruple

$$(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{G}) \equiv \left(\hat{A}_r, \begin{bmatrix} \Delta_u^T & \Delta_y^T \end{bmatrix}^T, \mathbf{I}_l^T, \begin{bmatrix} \Delta \mathbf{D}^T & 0 \end{bmatrix}^T \right).$$

The process and measurement noise are given by $\mathcal{M}_1 \mathbf{W}(k-i) + \mathcal{M}_2 \mathbf{V}(k-i)$ and $\mathbf{E} \mathbf{W}(k-i) + \mathbf{V}(k-i)$, respectively. The matrices \mathcal{A} and \mathcal{C} and the order of this system are already known. Given that $\Delta f^s(k) = -\hat{f}^s(k-i)$ corresponding to the second segment of the healthy data, and \mathcal{A} and \mathcal{C} , one can *estimate* \mathcal{B} and $\mathcal{G} = \begin{bmatrix} \mathcal{G}_1 & 0 \end{bmatrix}$ by invoking an optimization problem that is described in detail below.

Let us define the matrices $\mathcal{E}_{1,j}(k)$ and $\mathcal{Z}_{(\lambda-1),j}(k)$ similar to $\mathbf{G}_{i,j}(k)$ by replacing the signal $g(k)$ with the signals $-\hat{f}^s(k-i)$ and $\mathbf{Z}(k)$, respectively, where $\mathbf{Z}(k) = \begin{bmatrix} \mathbf{U}(k) \\ \mathbf{Y}(k) \end{bmatrix}$, and j is selected as large as the available data of the second segment allows. Therefore, for the system (47), we have,

$$\mathcal{E}_{1,j}(k) = \mathcal{C} \mathcal{A}^\lambda \xi(k-\lambda) + \mathcal{T}_{1,\lambda} \mathcal{Z}_{(\lambda-1),j}(k-\lambda) + ST \quad (48)$$

where the term ST denotes the stochastic terms which have zero mean and are neglected here for sake of brevity, and $\mathcal{T}_{1,\lambda}$ is defined as,

$$\mathcal{T}_{1,\lambda} = \begin{pmatrix} \mathcal{H}_{\lambda-1} & \mathcal{H}_{\lambda-2} & \dots & \mathcal{H}_0 & \mathcal{G} \end{pmatrix} \quad (49)$$

where $\mathcal{H}_\beta = \mathcal{C} \mathcal{A}^\beta \mathcal{B}$. The definition (49) shows dependence of $\mathcal{T}_{1,\lambda}$ on the matrices \mathcal{B} and \mathcal{G} . If λ is selected such that $\mathcal{A}^\lambda \approx 0$, then according to equation (48), one can obtain the estimates $\hat{\mathcal{B}}$ and $\hat{\mathcal{G}}$ by invoking the following minimization problem,

$$\begin{aligned} & \underset{\mathcal{B}, \mathcal{G}_1}{\text{minimize}} && \|\mathcal{E}_{1,j}(k) - \mathcal{T}_{1,\lambda} \mathcal{Z}_{(\lambda-1),j}(k-\lambda)\|_2 \\ & \text{subject to} && (\mathcal{B}(:, im+1 : im+il))^{p+} = 0. \end{aligned} \quad (50)$$

The constraint above does in fact enforce the columns p, \dots, ip of Δ_y to be equal to zero. For the case of sensor fault estimation problem, \hat{A}_r is selected to be an arbitrary Hurwitz matrix, therefore $\hat{A}_r = A_r$. On the other hand, $\Delta_y = \delta A = \hat{A}_r - \mathbf{M} + \hat{L}_r^{p\downarrow} = A_r - \mathbf{M} + \hat{L}_r^{p\downarrow} = -L_r^{p\downarrow} + \hat{L}_r^{p\downarrow}$. Therefore, the columns p, \dots, ip corresponding to Δ_y should be equal to zero.

A methodology for solving the minimization problem (50) is provided in the Appendix F. The solution will be consistent if the matrix $\mathcal{Z}_{(\lambda-1),j}(k)$ is full row rank which is not generally guaranteed. This condition on the matrix $\mathcal{Z}_{(\lambda-1),j}(k)$ depends on the nature of the system feedback control, the excitation signal and the available data and the real model of the system ([16,21,26]).

If the above condition on $\mathcal{Z}_{(\lambda-1),j}(k)$ is not satisfied, then the solution that is obtained by invoking the pseudo-inverse of $\mathcal{Z}_{(\lambda-1),j}(k)$ still minimizes the cost function in (50), but it will be biased.

An estimate of the error dynamics in presence of the sensor faults is therefore given by,

$$\begin{cases} \hat{\xi}(k+1) = \hat{A}_r \hat{\xi}(k) + \hat{\mathcal{B}} \mathbf{Z}(k-i) - \hat{\Delta}_y \mathbf{F}^s(k-i) \\ \Delta \hat{f}^s(k) = \mathbf{I}_l^T (\hat{\xi}(k) + \hat{\mathcal{G}}_1 \mathbf{U}(k-i)) \end{cases} \quad (51)$$

where $\Delta \hat{f}^s(k)$ is an estimate of $\Delta f^s(k)$. We assumed that the sensors $\sim p$ are healthy for the purpose of fault estimation of the sensors p . Moreover, the columns p, \dots, ip of $\hat{\Delta}_y$ are enforced to be zero in the minimization problem (50). Therefore, $\hat{\Delta}_y \mathbf{F}^s(k-i)$ is practically zero.

Consequently, one can now construct a new and a so-called *tuned* sensor fault estimation filter that is governed by,

$$\begin{cases} \eta(k+1) = \hat{A}_r \eta(k) + \tilde{\mathcal{B}}_s \mathbf{Z}(k-i) \\ \hat{f}^s(k-i) = \mathbf{I}_l^T (\eta(k) - \mathbf{Y}(k-i) + \tilde{\mathcal{D}}_s \mathbf{U}(k-i)) \end{cases} \quad (52)$$

where $\hat{f}^s(k-i)$ denotes the tuned estimate of the sensor fault, $\tilde{\mathcal{B}}_s = \begin{bmatrix} \hat{B}_r & \hat{L}^{p\downarrow} \end{bmatrix} + \hat{\mathcal{B}}$ and $\tilde{\mathcal{D}}_s = \hat{\mathbf{D}} + \hat{\mathcal{G}}_1$, where $\hat{\mathcal{B}}$ and $\hat{\mathcal{G}}_1$ are obtained and given by the minimization problem (50).

5.2 Actuator Fault Estimation Filters

The same procedure can now be followed for the actuator fault estimation filter. First, the following fault filter estimation is considered for the actuators q by using the data from the sensors $\sim p$ and actuators $\sim q$, namely

$$\begin{cases} \eta(k+1) = \hat{A}_r \eta(k) + \hat{B}_r \mathbf{U}^{q-}(k-i) + \hat{L}_r \mathbf{Y}^{p-}(k-i) \\ \hat{f}^a(k-i) = -\mathbf{I}_m^m \hat{\mathbf{D}}_{p-}^\dagger (\eta(k) - \mathbf{Y}^{p-}(k-i) + \hat{\mathbf{D}}_{p-} \mathbf{U}(k-i)) \end{cases} \quad (53)$$

where $\hat{f}^a(k-i)$ denotes an estimate of $f^a(k-i)$ and the filter parameters are obtained from the solution to the equations (34) to (37).

However, following along the same lines as those used in the Subsection 5.1, the so-called *tuned* actuator fault

estimation filter is now proposed as follows,

$$\begin{cases} \eta(k+1) = \hat{A}_r \eta(k) + \tilde{B}_a \mathbf{Z}^{q-,p-}(k-i) \\ \hat{f}^a(k-i) = -\mathbf{I}_m^m \tilde{D}_a^\dagger (\eta(k) - \mathbf{Y}^{p-}(k-i) + \tilde{D}_a \mathbf{U}(k-i)) \end{cases} \quad (54)$$

where $\tilde{B}_a = \begin{bmatrix} \hat{B}_r & \hat{L}_r \end{bmatrix} + \hat{B}$ and $\tilde{D}_a = \hat{D}_{p-} + \hat{G}_1$. The parameters \hat{B} and \hat{G}_1 are obtained by invoking the following minimization problem,

$$\underset{\mathcal{B}, \mathcal{G}_1}{\text{minimize}} \quad \|\mathcal{E}_{1,j}(k) - \mathcal{T}_{1,\lambda} \mathbf{Z}_{(\lambda-1),j}^{q-,p-}(k-\lambda)\|_2 \quad (55)$$

where the matrices $\mathcal{E}_{1,j}(k)$ and $\mathbf{Z}_{(\lambda-1),j}^{q-,p-}(k)$ are constructed similar to $\hat{\mathbf{G}}_{i,j}(k)$ by replacing the signal $g(k)$ with the signals $-\hat{f}^a(k-i)$ and $\mathbf{Z}^{q-,p-}(k)$, respectively,

and where $\mathbf{Z}^{q-,p-}(k) = \begin{bmatrix} \mathbf{U}^{q-}(k) \\ \mathbf{Y}^{p-}(k) \end{bmatrix}$, and j is selected as

large as the available data corresponding to the second segment allows. Note that the signal $-\hat{f}^a(k-i)$ for construction of the matrix $\mathcal{E}_{1,j}(k)$ is obtained by stimulating the filter (53) with the second segment of the healthy I/O data using an arbitrary initial condition for $\eta(0)$.

The above minimization problem is solved similar to the problem (50) as described in the Appendix F. We will demonstrate in the next section that the above tuning procedures will significantly improve the faults estimation accuracy performance. Similarly, the same tuning procedure can be applied to the fault detection and isolation filters in order to improve their performance when applied to a specific application. However, these details are left as topics of our future work.

Remark 6 As stated earlier in this section, one should partition the available off-line I/O data before applying the above tuning procedure. The length of the data in the segment that is used for solving the minimization problem (50) or (55) must be at least greater than λ , where λ was selected such that $\mathcal{A}^\lambda \approx 0$.

6 Simulation Results

In this section, we provide two numerical examples and simulations to illustrate the merits and advantages of our proposed schemes. In both cases, the healthy input is generated by a Pseudo Random Binary Signal (PRBS) generator. The system healthy output is generated by simulating it subject to healthy input in addition to state and measurement noise ($\mathcal{N}(0, 0.1)$) as governed by the dynamics \mathbf{S} . The Markov parameters are estimated by using the MATLAB built-in function *impzest*.

Fault Detection and Isolation Results: We consider

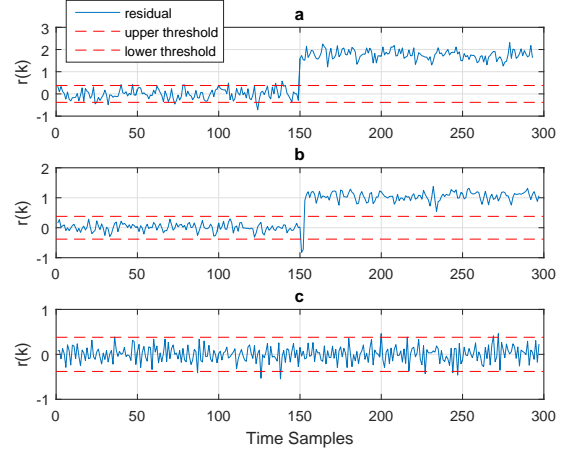


Fig. 1. A fault is injected in the actuator 1 of the system (56) at $k = 150$. (a) The output of the residual generator filter for achieving the fault detection task, (b) The output of the residual generator filter insensitive to the fault in the actuator 2, and (c) The output of the residual generator filter that is insensitive to the fault in the actuator 1.

the following non-minimum phase system which includes the fault model for the actuator bias (f_k^a) and sensor bias (f_k^s) as additive terms,

$$\begin{aligned} x_{k+1} &= \begin{bmatrix} 0 & 0 & 0 & -0.01 \\ 1 & 0 & 0 & 0.08 \\ 0 & 1 & 0 & -0.27 \\ 0 & 0 & 1 & -0.54 \end{bmatrix} x_k + \begin{bmatrix} 1 & -0.3 \\ 0 & 3.82 \\ 0 & 1.55 \\ 0 & -0.61 \end{bmatrix} (u_k + f_k^a) \\ y_k &= \begin{bmatrix} 1.58 & 0.725 & -0.60 & 0.31 \\ 2.4 & -0.08 & 0.42 & -0.05 \end{bmatrix} x_k + f_k^s \end{aligned} \quad (56)$$

The poles and zeros of the above system are located at $\{-0.39 \pm 53j, 0.11 \pm 0.09j\}$ and $\{0.17, 1.49\}$, respectively. Figure 1a shows the output of the residual generator filter (equation (38)) for performing the fault detection task by setting $p = q = \{\emptyset\}$ when a bias fault is injected in the actuator 1 at the time instant $k = 150$. We set $i = 2$. The identification data include 1000 samples. The numerical values for the detection filter are provided in the Appendix G. Figures 1b and 1c depict the outputs of the fault isolation filters 1 and 2 having the setting $q = \{1\}, p = \{\emptyset\}$ and $q = \{2\}, p = \{\emptyset\}$, respectively, and $i = 2$ for both. We have not yet applied the tuning process that was discussed in Section 5 to the above results, nevertheless these results demonstrate that actuator faults are successfully detected and isolated by application of our proposed data-driven methodology. In the next example, we will demonstrate the effects of the filter tuning process on the performance of the fault estimation accuracy.

Fault Estimation Results: Consider now the following minimum phase system,

$$\begin{aligned}
x_{k+1} &= \begin{bmatrix} -0.05 & -0.40 & 0 & -0.08 \\ -0.29 & -0.11 & 0.05 & -0.03 \\ -0.06 & 0.18 & -0.43 & 0.36 \\ 0.28 & 0.18 & -0.43 & 0.36 \end{bmatrix} x_k \\
&+ \begin{bmatrix} -0.15 & -0.99 \\ 0 & 0 \\ -0.68 & 0.07 \\ -0.96 & -0.20 \end{bmatrix} (u_k + f_k^a) \\
y_k &= \begin{bmatrix} -2.08 & 0 & -0.69 & 0 \\ 0 & -0.84 & 0.20 & 0.89 \end{bmatrix} x_k + f_k^s \quad (57)
\end{aligned}$$

The poles and zeros of the system are located at $\{-0.37, 0.30, -0.51 \pm 0.52j\}$ and $\{0.08, -0.58\}$, respectively. We next present a typical simulation result for estimating a fault in the system (57), and then provide comprehensive Monte Carlo simulations. Assume that a fault having a severity of 2 is injected in the sensor 2 at the time step $k = 150$. We selected a relatively large amplitude input signal given below to magnify the presence of biases,

$$u(k) = \begin{bmatrix} 20 + 20 \sin(5k) \\ 30 + 30 \cos(7k) \end{bmatrix} \quad (58)$$

We set $i = 2$, $p = \{2\}$ and $q = \{\emptyset\}$. We used 700 data samples for estimation of the Markov parameters and $\hat{\mathbf{M}}$ (equation (28)) and 300 samples for the filter tuning process. First, we tested the performance of our proposed sensor fault estimator (43) that is shown in Figure 2(a). The results clearly indicate that the filter is biased with an estimation error of 24%. The numerical values of the filter matrices are given in the Appendix H. We then tune the filter by solving the minimization problem (50) and construct the estimation filter as described by equation (52).

The numerical values for the matrices of the tuned filter are given in the Appendix H. The resulting estimation error for the tuned filter is now 1% as shown in Figure 2(b) which illustrates a significantly improved and enhanced performance as compared to those depicted in Figure 2(a). A better illustration of the improved performance is now provided through Monte Carlo simulation runs as described below. **Monte Carlo Simulation Results:** We have conducted Monte Carlo simulation runs for estimation of the faults in the system (57). We set $i = 2$, $p = \{1, 2\}$ and $q = \{\emptyset\}$ and $i = 2$, $p = \{\emptyset\}$ and $q = \{1, 2\}$ for the sensor and the actuator fault estima-

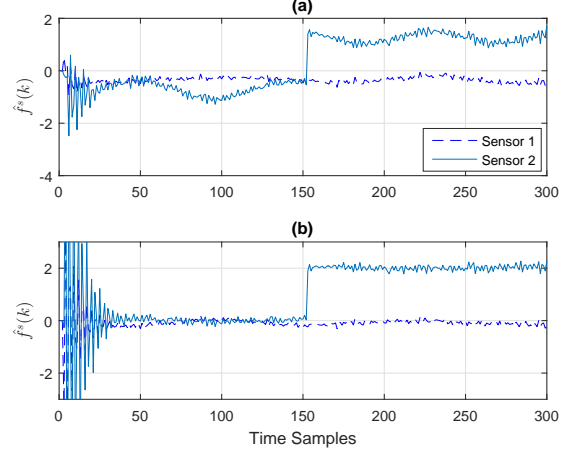


Fig. 2. A fault having a severity of 2 is injected in the sensor 2 of the system (57) at $k = 150$. (a) The output of the original sensor fault estimation filter, and (b) The output of the tuned sensor fault estimation filter.

Table 1

The Monte Carlo simulation results for estimation of the faults in the system (57) using different filters and under two different system inputs, where μ and σ denote the mean and variance, respectively. The filters are specifically the sensor fault estimator (43), the tuned sensor fault estimator (52), the actuator fault estimator (53) and the tuned fault estimator (54) denoted by F(43), F(52), F(53) and F(54), respectively.

	$u_1(k)$		$u_2(k)$	
	$\mu(\Delta f)$	$\sigma(\Delta f)$	$\mu(\Delta f)$	$\sigma(\Delta f)$
F(43)	(-0.47, -0.06)	(0.30, 1.13)	(-0.03, 0)	(0.0, 0.02)
F(52)	(-0.02, 0)	(0.07, 0.18)	(-0.04, 0)	(0, 0.01)
F(53)	(1.31, -4.02)	(3.2, 11.3)	(0.14, -0.4)	(0.02, 0.09)
F(54)	(-0.01, 0.01)	(0.01, 0.02)	(0, 0)	(0, 0)

tions, respectively. Sensors faults having severities of -1 and 1 are injected to the sensor 1 and the sensor 2 at the time step $k = 150$, respectively. The same fault scenario is considered for the actuator fault estimation problem. We performed 400 Monte Carlo simulation runs for two inputs that are selected as $u_1(k) = u(k)$ given by equation (58) and $u_2(k) = 0.1u(k)$. The results are shown in Figure 3 and numerically presented in Table 1. It can be concluded that the filters (43) and (53) have an acceptable performance for relatively small inputs (in terms of the norm of the signal). On the other hand, relatively large inputs clearly magnify the biases although they are well-managed by utilizing our proposed tuning process. An approximation to the biases for the filters (43) and (53) can be obtained by using equation (46). The L_2 gain

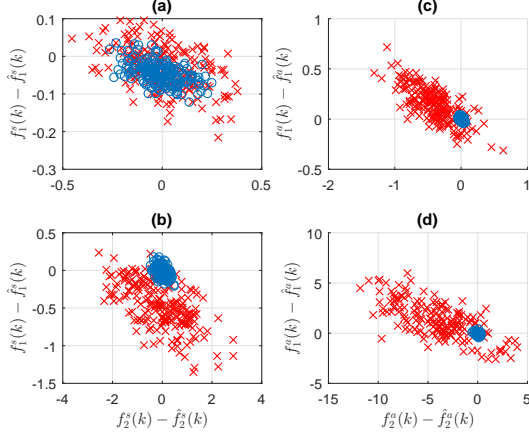


Fig. 3. Faults having severities of 1 and -1 are injected in the actuators or sensors of the system (57). The red marks represent results for the filters (43) or (53) and the blue circles represent results for the tuned filters (52) or (54). (a) Sensors fault estimation error when the system is stimulated by $u_2(k)$, (b) Sensors fault estimation error when the system is stimulated by $u_1(k)$, (c) Actuator fault estimation errors when the system is stimulated by $u_2(k)$, and (d) Actuator fault estimation errors when the system is stimulated by $u_1(k)$.

of the error dynamics is then given by,

$$\|\Delta f^s(k)\|_2 \leq \|(zI - \hat{A}_r)\hat{\mathcal{B}} + \hat{\mathcal{G}}\|_\infty \|\mathbf{Z}(k)\|_2$$

The matrices $\hat{\mathcal{B}}$ and $\hat{\mathcal{G}}_1$ are obtained by solving the minimization problems (50) and (55). Therefore, one can obtain a prediction of the error margin corresponding to a certain input.

Comparative Study: Finally, in order to perform a comparative study to demonstrate the capability and advantage of our proposed methodology, we consider the example that was provided in [16] and evaluate our corresponding results with those in [16]. The system in [16] is a continuous-time system and represents a linearized model of a vertical take-off and landing (VTOL) aircraft

that is given by,

$$\begin{aligned} \dot{x}(t) &= \begin{bmatrix} -0.036 & 0.027 & 0.018 & -0.455 \\ 0.048 & -1.01 & 0.002 & -4.020 \\ 0.100 & 0.368 & -0.707 & 1.42 \\ 0 & 0 & 1 & 0 \end{bmatrix} x(t) \\ &+ \begin{bmatrix} 0.44 & 0.17 \\ 3.54 & -7.59 \\ -5.52 & 4.49 \\ 0 & 0 \end{bmatrix} (u(t) + f^a(t)) \\ y(t) &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix} x(t) + f_s(t) \end{aligned} \quad (59)$$

where $f^s(t) \in \mathbb{R}^4$ and $f^a(t) \in \mathbb{R}^2$ and $f^a(t)$ with $f^s(t)$ representing the actuator and sensor bias faults, respectively. The discrete-time model associated with the system (59) is obtained by using a sampling rate of 0.5 seconds. Furthermore, it is assumed that the system is stabilized by applying the following control law which is experimentally obtained,

$$u(k) = - \begin{bmatrix} 0 & 0 & -0.5 & 0 \\ 0 & 0 & -0.1 & -0.1 \end{bmatrix} y(k) + \xi(k)$$

where $\xi(k)$ denotes the reference signal. The process and measurement noise are white having zero mean and covariances $Q = 0.16\mathbf{I}$ and $R = 0.64\mathbf{I}$, respectively. The reference signal is selected to be a PRBS signal for identification of the Markov parameters.

The identification data set includes 1000 samples. Note that the *correlation analysis* cannot be directly applied to unstable systems. Therefore, response of the stable closed-loop system is obtained by injecting the input that is computed at each time step using the above control law. Next, the input and closed-loop system responses are used as I/O data for identification of the closed-loop system Markov parameters through the *correlation analysis*. The injected fault signals to the actuators and sensors are given by,

$$f^a(k) = \begin{cases} \begin{bmatrix} 0 & 0 \end{bmatrix}^T & 0 \leq k \leq 50 \\ \begin{bmatrix} \sin(0.1\pi k) & 1 \end{bmatrix}^T & k > 50 \end{cases}$$

$$f^s(k) = \begin{cases} \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}^T & 0 \leq k \leq 50 \\ \begin{bmatrix} \sin(0.1\pi k) & 1 & 0 & 0 \end{bmatrix}^T & k > 50 \end{cases}$$

We have set $i = 2$, $p = \{1, 2\}$ and $q = \{\emptyset\}$ and $i = 3$, $p = \{\emptyset\}$ and $q = \{1, 2\}$ for the sensor and actuator fault estimation filters, respectively. The reference signal is set to $\xi(k) = 15$ in order to duplicate the Monte Carlo simulation results that were reported in [16].

Figures 4a and 4b show 500 and 400 Monte Carlo simulation runs for estimation of the actuator and sensor faults, respectively. The average estimation errors are given by $\mu(\hat{f}_1^a - f_1^a, \hat{f}_2^a - f_2^a) = (0.018, 0.039)$ and $\mu(\hat{f}_1^s - f_1^s, \hat{f}_2^s - f_2^s) = (0.005, 0.139)$. The variances are given by $\sigma(\hat{f}_1^a - f_1^a, \hat{f}_2^a - f_2^a) = (0.008, 0.0097)$ and $\sigma(\hat{f}_1^s - f_1^s, \hat{f}_2^s - f_2^s) = (0.0398, 0.1074)$.

The above results clearly show that our proposed scheme has significant advantages, benefits, and capabilities over the receding horizon fault estimator that was proposed in [16], although it uses the same set of assumptions. This is substantiated by the following observations. First, our proposed filter order is significantly lower than that in [16] as we have theoretically shown in Theorems 1 and 2. For this particular example, we have used $i = 2$ for the sensor and actuator fault estimation filters, respectively, whereas i is set to $i = 30$ in [16]. Moreover, we have achieved a better performance by invoking an offline tuning procedure as compared to the Algorithm 3 utilized in [16] that performs an online optimization solution. Consequently, the computational burden of [16] to the user increases to the point where the average computational time per sample takes 2.05 seconds on a 3.4 GHz computer having 8 GB of RAM. Whereas, the computational time associated with our proposed methodology per sample using the same computer takes only 8.2×10^{-7} seconds.

7 Conclusion

We have proposed fault detection, isolation and estimation schemes that are all directly constructed and designed in the state-space representation form by utilizing only the system I/O data. We have shown that to design and develop our schemes it is only sufficient to estimate the system Markov parameters. Consequently, the reduction step that is commonly used in the literature, and that also introduces nonlinear errors, and also requires an *a priori* knowledge of the system order is completely eliminated in our schemes. We have shown that the performance of the estimation scheme is linearly dependent on the Markov parameters estimation process errors. We also proposed an offline tuning procedure that effectively compensates for the estimation errors that are caused by errors in the estimation of the Markov parameters. Comparisons of our proposed schemes with those avail-

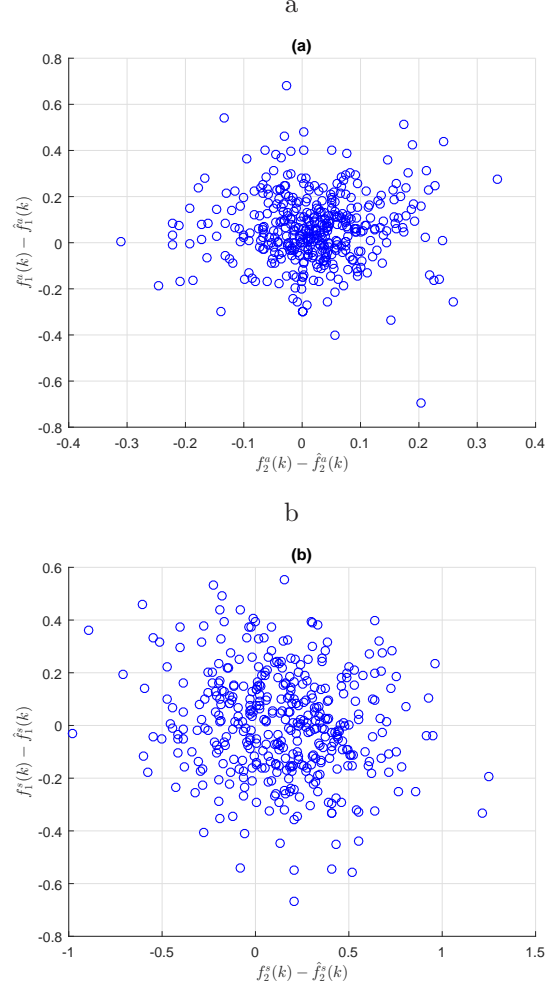


Fig. 4. (a) The first actuator fault estimation error versus the second actuator fault estimation error for the system (59) using 500 Monte Carlo simulation runs, and (b) The first sensor fault estimation error versus the second sensor fault estimation error for the system (59) using 400 Monte Carlo simulation runs.

able in the literature have revealed that our methodology is mathematically simpler to develop and computationally more efficient, while it maintains the same level of performance and requires a lower set of assumptions. Further research is required to investigate the robustness of our scheme to estimation errors and presence of concurrent sensor and actuator faults.

References

- [1] R. Beard, "Failure accommodation in linear systems through self-reorganization," *Ph.D dissertation, Dept. Aeronautics and Astronautics, MIT*, 1971.
- [2] Z. Gao, C. Cecati, and S. Ding, "A survey of fault diagnosis and fault-tolerant techniques; part i: Fault diagnosis with model-based and signal-based approaches," *IEEE Transactions on Industrial Electronics*, vol. 62, pp. 3757–3767, June 2015.

[3] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques; part ii: Fault diagnosis with knowledge-based and hybrid/active approaches," *IEEE Transactions on Industrial Electronics*, vol. 62, pp. 3768–3774, June 2015.

[4] I. Hwang, S. Kim, Y. Kim, and C. E. Seah, "A survey of fault detection, isolation, and reconfiguration methods," *IEEE Transactions on Control Systems Technology*, vol. 18, no. 3, pp. 636–653, 2010.

[5] S. S. Tayarani-Bathaie and K. Khorasani, "Fault detection and isolation of gas turbine engines using a bank of neural networks," *Journal of Process Control*, vol. 36, pp. 22 – 41, 2015.

[6] A. Rodríguez Ramos, C. Domínguez Acosta, P. J. Rivera Torres, E. I. Serrano Mercado, G. Beauchamp Baez, L. A. Rifón, and O. Llanes-Santiago, "An approach to multiple fault diagnosis using fuzzy logic," *Journal of Intelligent Manufacturing*, pp. 1–11, 2016.

[7] E. Sobhani-Tehrani, H. Talebi, and K. Khorasani, "Hybrid fault diagnosis of nonlinear systems using neural parameter estimators," *Neural Networks*, vol. 50, pp. 12 – 32, 2014.

[8] S. Ding, P. Zhang, A. Naik, E. Ding, and B. Huang, "Subspace method aided data-driven design of fault detection and isolation systems," *Journal of Process Control*, vol. 19, no. 9, pp. 1496–1510, 2009.

[9] J. Dong, B. Kulcsr, and M. Verhaegen, "Subspace based fault detection and identification for LPV systems," vol. 42, pp. 336 – 341, 2009. 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes.

[10] Y. Wang, G. Ma, S. X. Ding, and C. Li, "Subspace aided data-driven design of robust fault detection and isolation systems," *Automatica*, vol. 47, no. 11, pp. 2474–2480, 2011.

[11] S. Ding, "Data-driven design of monitoring and diagnosis systems for dynamic processes: A review of subspace technique based schemes and some recent results," *Journal of Process Control*, vol. 24, no. 2, pp. 431–449, 2014.

[12] J. Dong and M. Verhaegen, "Subspace based fault detection and identification for LTI systems," *IFAC Proceedings Volumes*, vol. 42, no. 8, pp. 330 – 335, 2009. 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes.

[13] J. Dong, M. Verhaegen, and F. Gustafsson, "Robust fault detection with statistical uncertainty in identified parameters," *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5064–5076, 2012.

[14] J. Dong, M. Verhaegen, and F. Gustafsson, "Robust fault isolation with statistical uncertainty in identified parameters," *IEEE Transactions on Signal Processing*, vol. 60, pp. 5556–5561, Oct 2012.

[15] J. Dong and M. Verhaegen, "Identification of fault estimation filter from i/o data for systems with stable inversion," *IEEE Transactions on Automatic Control*, vol. 57, no. 6, pp. 1347–1361, 2012.

[16] Y. Wan, T. Keviczky, M. Verhaegen, and F. Gustafsson, "Data-driven robust receding horizon fault estimation," *Automatica*, vol. 71, pp. 210 – 221, 2016.

[17] L. Ljung, *System identification*. Springer, 1998.

[18] P. Van Overschee and B. De Moor, "A unifying theorem for three subspace system identification algorithms," *Automatica*, vol. 31, no. 12, pp. 1853–1864, 1995.

[19] T. Katayama, *Subspace methods for system identification*. Springer Science & Business Media, 2006.

[20] B. Huang and R. Kadali, *Dynamic modeling, predictive control and performance monitoring: A data-driven subspace approach*, vol. 374. Springer Science & Business Media, 2008.

[21] A. Chiuso, "On the relation between cca and predictor-based subspace identification," *IEEE Transactions on Automatic Control*, vol. 52, no. 10, pp. 1795–1812, 2007.

[22] R. J. Patton, P. M. Frank, and R. N. Clark, *Issues of fault diagnosis for dynamic systems*. Springer Science & Business Media, 2013.

[23] S. Ding, *Model-based fault diagnosis techniques: design schemes, algorithms, and tools*. Springer Science & Business Media, 2008.

[24] S. Kirtikar, H. Palanthandalam-Madapusi, E. Zattoni, and D. S. Bernstein, "l-delay input reconstruction for discrete-time linear systems," in *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*, pp. 1848–1853, IEEE, 2009.

[25] S. Sundaram and C. N. Hadjicostis, "Delayed observers for linear systems with unknown inputs," *IEEE Transactions on Automatic Control*, vol. 52, pp. 334–339, Feb 2007.

[26] G. van der Veen, J.-W. Van Wingerden, M. Bergamasco, M. Lovera, and M. Verhaegen, "Closed-loop subspace identification methods: an overview," *IET Control Theory & Applications*, vol. 7, no. 10, pp. 1339–1358, 2013.

A Proof of Theorem 1

The dynamics of the residuals in presence of actuator faults is given by,

$$\begin{cases} \eta(k+1) = A_r \eta(k) + B_r^{q-} (\mathbf{U}^{q-}(k-i) + \mathbf{F}^{a,q-}(k-i)) \\ L_r^{p-} \mathbf{Y}^{p-}(k-i) \\ r_q(k) = \eta(k) - \mathbf{Y}^{p-}(k-i) \\ + \mathbf{D}_{p-} (\mathbf{U}(k-i) + \mathbf{F}^a(k-i)) \end{cases} \quad (\text{A.1})$$

Since A_r , B_r and L_r satisfy equations (14) and (15) to (18) and the actuators $\sim q$ are healthy, therefore $\mathbb{E}\{\eta(k) - \mathbf{C}_{p-}x(k-i)\} \rightarrow 0$ as $k \rightarrow \infty$. The convergence is asymptotic since A_r is Hurwitz. Therefore, $\mathbb{E}\{r_q(k)\} \rightarrow -\mathbf{D}_{p-}\mathbf{F}^a(k-i)$ as $k \rightarrow \infty$. Therefore, $\hat{f}^a(k-i) = -\mathbf{I}_l^m \mathbf{D}^\dagger \mathbb{E}\{r_q(k)\} - \mathbf{I}_m^m (\mathbf{I} - \mathbf{D}_{p-}^\dagger \mathbf{D}_{p-}) \Theta$, where Θ is an arbitrary matrix. However, since $i \geq \tau_p$, the subspace spanned by rows of \mathbf{I}_m^m are also spanned by the rows of \mathbf{D}_{p-} . Therefore, the projection of the row space of \mathbf{I}_m^m onto the null space of \mathbf{D}_{p-} given by $(\mathbf{I} - \mathbf{D}_{p-}^\dagger \mathbf{D}_{p-})$ is zero. In other words, $\mathbf{I}_m^m (\mathbf{I} - \mathbf{D}_{p-}^\dagger \mathbf{D}_{p-}) = 0$. This completes the proof of the theorem.

B Proof of Theorem 2

The dynamics of the residuals in presence of sensor faults is given by,

$$\begin{cases} \eta(k+1) = A_r \eta(k) + B_r \mathbf{U}(k-i) \\ + L_r^{p\downarrow} (\mathbf{Y}(k-i) + \mathbf{F}^s(k-i)) \\ r_p(k) = \eta(k) - \mathbf{Y}(k-i) - \mathbf{F}^s(k-i) + \mathbf{D}_{p-} \mathbf{U}(k-i) \end{cases} \quad (\text{B.1})$$

Since A_r , B_r and $L_r^{p\downarrow}$ satisfy equations (14) and (15) to (18) and the actuators and sensors p are healthy, therefore $\mathbb{E}\{\eta(k) - \mathbf{C}_{p-}x(k-i)\} \rightarrow 0$ as $k \rightarrow \infty$. The convergence is asymptotic since A_r is Hurwitz. Therefore, $\mathbb{E}\{r_p(k)\} \rightarrow \mathbf{F}^s(k-i)$ as $k \rightarrow \infty$. Therefore, $\hat{f}^s(k-i) = \mathbf{I}_l^m \mathbb{E}\{r_p(k)\}$. This completes the proof of the theorem.

C Proof of Lemma 1

According to equation (26) and the measurement equation (4), we have

$$\Gamma_0^{p-} = \mathbf{C}_{p-}X(k-i) + \mathbf{E}_{p-}\mathbf{W}_{i,j}(k-i) + \mathbf{V}_{i,j}^{p-}(k-i)$$

where,

$$X(k-i) = \begin{bmatrix} x(k-i) & \dots & x(k-i+j-1) \end{bmatrix}$$

Assume that two block rows γ and β of Γ_0^{p-} , where $\beta > \gamma$, are linearly dependent which implies,

$$\begin{aligned} & \mathbf{Y}_{1,j}^{p-}(k-i+\gamma) \\ & - (\mathbf{D}_{p-}((i+\gamma-1)l' : (i+\gamma)l', :))\mathbf{U}_{i,j}(k-i) \\ & = c(\mathbf{Y}_{1,j}^{p-}(k-i+\beta) \\ & - (\mathbf{D}_{p-}((i+\beta-1)l' : (i+\beta)l', :))\mathbf{U}_{i,j}(k-i)) \end{aligned} \quad (\text{C.1})$$

where c is a constant. Equivalently, we have,

$$\begin{aligned} & CA^{\gamma-1}X(k-i) \\ & + (\mathbf{E}_{p-}((i+\gamma-1)l' : (i+\gamma)l', :))\mathbf{W}_{i,j}(k-i) \\ & + \mathbf{V}_{1,j}^{p-}(k-i+\gamma) = c(CA^{\beta-1}X(k-i) \\ & + (\mathbf{E}_{p-}((i+\beta-1)l' : (i+\beta)l', :))\mathbf{W}_{i,j}(k-i) + \mathbf{V}_{1,j}^{p-}(k-i+\beta)) \end{aligned} \quad (\text{C.2})$$

If we multiply both sides of equation (C.2) by $(\mathbf{V}_{1,j}^{p-}(k-i+\beta))^T$ and take the limit as $j \rightarrow \infty$, all the terms will be zero except for the last one since all the terms except the last one are uncorrelated with $(\mathbf{V}_{1,j}^{p-}(k-i+\beta))$. Therefore, we obtain $0 = c$, which is a contradiction. Therefore, $\hat{\Gamma}_0^{p-}$ is full row rank.

D Proof of Lemma 3

Let us define $\psi(k-i) = \mathbf{Y}_{i,j}^{p-}(k-i) - \hat{\mathbf{D}}_{p-}\mathbf{U}_{i,j}(k-i)$. Therefore, $\psi(k-i+1)$ is governed by,

$$\begin{aligned} \psi(k-i+1) &= \mathbf{Y}_{i,j}^{p-}(k-i+1) - \hat{\mathbf{D}}_{p-}\mathbf{U}_{i,j}(k-i+1) \\ &= \mathbf{Y}_{(i+1),j}^{p-}(k-i+1) - \hat{\mathbf{D}}_{+,p-}\mathbf{U}_{(i+1),j}(k-i) \\ &\quad - \hat{\mathcal{D}}_{+,p-}\mathbf{I}_{l'}^m\mathbf{U}_{i,j}(k-i) \\ &= \hat{\mathbf{M}}_{p-}\psi(k-i) - \hat{\mathcal{D}}_{+,p-}\mathbf{I}_{l'}^m\mathbf{U}_{i,j}(k-i) \end{aligned} \quad (\text{D.1})$$

Due to the fact that $\psi(k-i)$ is bounded, therefore $\hat{\mathbf{M}}_{p-}$ is a Hurwitz matrix. This completes the proof of the lemma.

E Proof of Lemma 4

We show that the first $(i-1)l$ rows of the residuals generated by the filter (38) approach to zero as $k \rightarrow \infty$

if $q = \{\emptyset\}$. We begin by noting that,

$$\hat{r}(k+1) = \eta(k+1) - \mathbf{Y}^{p-}(k-i+1) + \hat{\mathbf{D}}_{p-}\mathbf{U}(k-i+1) \quad (\text{E.1})$$

Substituting $\eta(k+1)$ from the state equation of the filter (38) yields,

$$\begin{aligned} \hat{r}(k+1) &= \hat{A}_r\eta(k) + \hat{B}_r\mathbf{U}(k-i) + \hat{L}_r\mathbf{Y}(k-i) \\ &\quad - \mathbf{Y}^{p-}(k-i+1) + \hat{\mathbf{D}}_{p-}\mathbf{U}(k-i+1) \end{aligned} \quad (\text{E.2})$$

Next, we substitute \hat{B}_r and \hat{L}_r from equations (35), (36) and (37) and rearrange it to obtain,

$$\begin{aligned} \hat{r}(k+1) &= \hat{A}_r r(k) - \hat{\mathbf{M}}_{p-} \left(\mathbf{Y}^{p-}(k-i) - \hat{\mathbf{D}}_{p-}\mathbf{U}(k-i) \right) \\ &\quad - \mathbf{Y}^{p-}(k-i+1) + \hat{\mathbf{D}}_{p-}\mathbf{U}(k-i+1) \\ &\quad + \hat{\mathcal{D}}_{+,p-}\mathbf{I}_{l'}^m\mathbf{U}(k-i) \end{aligned} \quad (\text{E.3})$$

Note that,

$$\hat{\mathbf{D}}_{p-}\mathbf{U}(k-i+1) + \hat{\mathcal{D}}_{+,p-}\mathbf{I}_{l'}^m\mathbf{U}(k-i) = \hat{\mathbf{D}}_{+,p-}\mathbf{U}_+(k-i)$$

Therefore, by considering the structure of $\hat{\mathbf{M}}_{p-}$ in Lemma 2, one can verify that (E.3) becomes,

$$\hat{r}(k+1) = \hat{A}_r\hat{r}(k) + \begin{bmatrix} 0_{(i-1)l \times 1} \\ y^{p-}(k+1) - \hat{\mathbf{D}}_{+,p-}\mathbf{U}_+(k-i) \end{bmatrix} \quad (\text{E.4})$$

which shows that if \hat{A}_r is a diagonal Hurwitz matrix, then the first $(i-1)l$ rows of $\hat{r}(k)$ approach to zero as $k \rightarrow \infty$. Note that we did not use the relation $\hat{\mathbf{M}}\hat{\Gamma}_0(\cdot, 1) = \hat{\Gamma}_1(\cdot, 1)$ that only holds for the healthy system. In that case, we would clearly obtain $\hat{r}(k+1) = \hat{A}_r\hat{r}(k) + 0$, which is a valid model. This completes the proof of the lemma.

F A Methodology for Solving the Minimization Problems (50) and (55)

The constraint in the optimization problem enforces that columns that the multiplied by the rows p, \dots, ip of $\mathbf{Y}(k)$ should be zero. Therefore, they can be simply removed by using $\mathcal{Z}^{\{\emptyset\},p-}(k)$ instead of $\mathcal{Z}(k)$ and by invoking now the following optimization problem,

$$\underset{\mathcal{B}^\phi, \mathcal{G}_1}{\text{minimize}} \quad \|\mathcal{E}_{1,j}(k) - \mathcal{T}_{1,\lambda}^\phi \mathcal{Z}_{(\lambda-1),j}^{\{\emptyset\},p-}(k-\lambda)\|_2 \quad (\text{F.1})$$

where

$$\mathcal{T}_{1,\lambda}^\phi = \left(CA^{\lambda-1}\mathcal{B}^\phi \quad CA^{\lambda-2}\mathcal{B}^\phi \quad \dots \quad C\mathcal{B}^\phi \quad \mathcal{G} \right) \quad (\text{F.2})$$

Once $\hat{\mathcal{B}}^\phi$ is computed, $\hat{\mathcal{B}}$ is easily constructed by inserting back zero columns at the columns $im + p, im + 2p, \dots, im + ip$ of the matrix $\hat{\mathcal{B}}^\phi$. An estimate of $\hat{\mathcal{T}}_{1,\lambda}^\phi$ is now given by,

$$\begin{aligned}\hat{\mathcal{T}}_{1,\lambda}^\phi &= \mathcal{E}_{1,j}(k) \left(\mathcal{Z}_{(\lambda-1),j}^{\{\emptyset\},p-}(k-s) \right)^\dagger \\ &= \begin{bmatrix} \hat{\mathcal{T}}_1 & \dots & \hat{\mathcal{T}}_\lambda \end{bmatrix}\end{aligned}\quad (\text{F.3})$$

where $\hat{\mathcal{T}}_\alpha \in \mathbb{R}^{l \times (im+il-in_p)}$. We reformulate the definition (F.2) in the matrix form as follows,

$$\begin{bmatrix} \hat{\mathcal{T}}_1 \\ \vdots \\ \hat{\mathcal{T}}_\lambda \end{bmatrix} = \begin{bmatrix} \mathfrak{C} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \hat{\mathcal{B}}^\phi \\ \hat{\mathcal{G}}_1 \end{bmatrix}\quad (\text{F.4})$$

where,

$$\mathfrak{C} = \begin{bmatrix} \mathcal{C} \\ \mathcal{C}\mathcal{A} \\ \vdots \\ \mathcal{C}\mathcal{A}^{\lambda-1} \end{bmatrix}\quad (\text{F.5})$$

Note that we have enforced $\mathcal{G} = \begin{bmatrix} \mathcal{G}_1 & 0 \end{bmatrix}$ in the right hand side of (F.4). The solution to (F.4) is now given by,

$$\begin{bmatrix} \hat{\mathcal{B}}^\phi \\ \hat{\mathcal{G}}_1 \end{bmatrix} = \begin{bmatrix} \mathfrak{C} & 0 \\ 0 & I \end{bmatrix}^\dagger \begin{bmatrix} \hat{\mathcal{T}}_1 \\ \vdots \\ \hat{\mathcal{T}}_\lambda \end{bmatrix}\quad (\text{F.6})$$

Note that the matrix $\begin{bmatrix} \mathfrak{C} & 0 \\ 0 & I \end{bmatrix}$ is full column rank. The above solution provides the least square solution to the minimization problem (F.1). However, if $\mathcal{Z}_{(s-1),j}^{\{\emptyset\},p-}(k-s)$ is full row rank, then the minimum value of zero will be achieved. For the problem (55), it is only sufficient to replace $\mathcal{Z}_{(\lambda-1),j}^{\{\emptyset\},p-}(k-\lambda)$, $\mathcal{T}_{1,\lambda}^\phi$ and $\hat{\mathcal{B}}^\phi$ in equations (F.3) and (F.1) by $\mathcal{Z}_{(\lambda-1),j}^{q-,p-}(k)$, $\mathcal{T}_{1,\lambda}$, and $\hat{\mathcal{B}}$, respectively. This provides the details on the methodology for solving the minimization problems (50) and (55).

G Numerical Values of the Matrices in Example 1

$$\hat{\mathbf{M}} = \begin{bmatrix} 0.00 & 0.00 & 1.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 1.00 \\ -0.30 & 0.12 & -0.51 & 0.51 \\ 0.28 & -0.11 & 0.31 & -0.16 \end{bmatrix}$$

$$\begin{aligned}\hat{A}_r &= \begin{bmatrix} 0.26 & 0 \\ 0 & 0.44 \end{bmatrix}; \hat{B}_r = \begin{bmatrix} 0.70 & -0.99 & 0 & 0 \\ 0.88 & 1.01 & 0 & 0 \end{bmatrix} \\ \hat{L}_r &= \begin{bmatrix} -0.30 & 0.12 & -0.78 & 0.5159 \\ 0.28 & -0.11 & 0.31 & -0.61 \end{bmatrix}\end{aligned}$$

H Numerical Values of the Matrices in Example 2

$$\hat{\mathbf{M}} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0.06 & -0.18 & -0.19 & -0.17 \\ -0.17 & -0.43 & -0.48 & -0.81 \end{bmatrix}$$

$$\hat{B}_r = \begin{bmatrix} 0.61 & 1.61 & 0 & 0 \\ -1.06 & -0.34 & 0 & 0 \\ -0.28 & -0.51 & 0 & 0 \\ -0.15 & -1.71 & 0 & 0 \end{bmatrix}; \hat{L}_r^{p\downarrow} = \begin{bmatrix} 0.12 & 0 & 0.20 & 0 \\ 0.59 & 0 & 0.08 & 0 \\ 0.24 & 0 & 0.19 & 0 \\ 0.28 & 0 & 0.59 & 0 \end{bmatrix}$$

$$\hat{A}_r = \hat{\mathbf{M}} - \hat{L}_r^{p\downarrow}; \hat{\mathbf{D}} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0.76 & 2.03 & 0 & 0 \\ -0.99 & -0.16 & 0 & 0 \end{bmatrix}$$

$$\hat{\mathcal{B}}(:, 1:4) = \begin{bmatrix} -0.07 & -0.03 & 0.07 & 0.14 \\ 0.01 & -0.16 & 0.03 & 0.09 \\ 0 & -0.09 & 0.03 & 0.06 \\ -0.04 & -0.16 & 0 & 0 \end{bmatrix}$$

$$\hat{\mathcal{B}}(:, 5:8) = \begin{bmatrix} 0.02 & 0 & 0 & 0 \\ -0.01 & 0 & 0.02 & 0 \\ -0.01 & 0 & 0.08 & 0 \\ -0.02 & 0 & 0.17 & 0 \end{bmatrix}; \hat{\mathcal{G}} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -0.05 & -0.07 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$